

# Prospective life tables

Heather Booth, Rob J Hyndman, Leonie Tickle

May 23, 2013

## 1 Introduction

Prospective life tables depend on forecasting age-specific mortality. Considerable attention has been paid to methods for forecasting mortality in recent years. Much of this work has grown out of the seminal Lee-Carter method (Lee & Carter 1992). Other extrapolative approaches use Bayesian modelling, generalized linear modelling and state-space approaches. Methods for forecasting mortality have been extensively reviewed by Booth (2006) and Booth & Tickle (2008). This chapter covers various extrapolative methods for forecasting age-specific central death rates. Also covered is the derivation of stochastic life expectancy forecasts based on mortality forecasts.

The main packages on CRAN for implementing life tables and mortality modelling are **demography** (Hyndman 2012) and **MortalitySmooth** (Camarda 2012) and we will concentrate on the methods implemented in those packages. However, mention is also made of other extrapolative approaches, and related R packages where these exist.

We will use, as a vehicle of illustration, US mortality data from 1950. This can be extracted from the Human Mortality Database (2013) using the **demography** package.

```
library(demography)
library(MortalitySmooth)
usa <- hmd.mx("USA", "username", "password", "USA")
usa1950 <- extract.years(usa, years=1950:2010)
```

The `username` and `password` are for the Human Mortality Database. In this chapter, we will assume that the above R commands have been entered.

## 2 Smoothing mortality data

Suppose  $D_{x,t}$  is the number of deaths in calendar year  $t$  of people aged  $x$ , and  $E_{x,t}^c$  is the total years of life lived between ages  $x$  and  $x+1$  in calendar year  $t$ , which can be approximated by the mid-year (central) population at age  $x$  in year  $t$ . Then the observed mortality rate is defined as

$$m_{x,t} = D_{x,t}/E_{x,t}^c.$$

Typically we observe deaths at single years of age ( $x_1 = 0, x_2 = 1, \dots$ ) or in 5-year age groups ( $x_1 = [0, 4], x_2 = [5, 9], \dots$ ).

In order to stabilize the high variance associated with high age-specific rates, it is necessary to transform the raw data by taking logarithms. Consequently, the mortality models considered in this chapter are all in log scale.

Figure 1 shows an example of such data for the USA. (Age-specific mortality rates can be higher than one for very small populations as the number of deaths of people aged  $x$  may exceed the mid-year population aged  $x$ .)

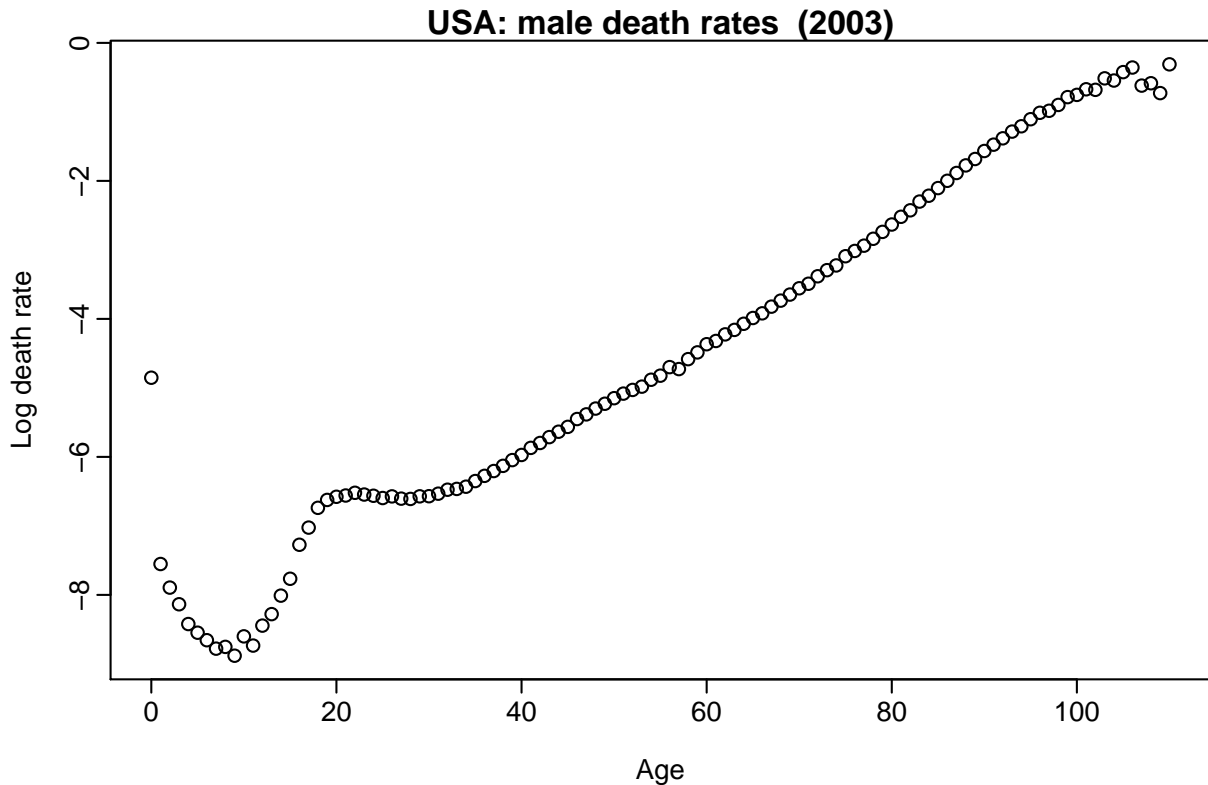


Figure 1: Male mortality rates for single years of age in the United States, 2003.

This example shows that the mortality rates follow a smooth function with some observational error. The observational error has higher variance at very old ages (when the populations are small) and at young ages (when the mortality rates are small).

Thus, we observe  $\{x_i, m_{x_i,t}\}$ ,  $t = 1, \dots, n$ ,  $i = 1, \dots, p$  where

$$\log m_{x_i,t} = f_t(x_i^*) + \sigma_t(x_i^*)\varepsilon_{t,i},$$

$\log$  denotes the natural logarithm,  $f_t(x)$  is a smooth function of  $x$ ,  $x_i^*$  is the mid-point of age interval  $x_i$ ,  $\varepsilon_{t,i}$  is an iid random variable and  $\sigma_t(x)$  allows the amount of noise to vary with  $x$ .

Then the observational variance,  $\sigma_t^2(x)$ , can be estimated assuming deaths are Poisson distributed (Brillinger 1986). Thus,  $m_{x,t}$  has approximate variance  $D_{x,t}/(E_{x,t}^c)^2$ , and the variance of  $\log m_{x,t}$  (via a Taylor approximation) is

$$\sigma_t^2(x) \approx 1/D_{x,t}.$$

Life tables constructed from the smoothed  $f_t(x)$  data have lower variance than tables constructed from the original  $m_{t,x}$  data, and thus provide better estimates of life expectancy. To estimate  $f$  we can use a nonparametric smoothing method such as kernel smoothing, loess, or splines. Two smoothing methods for estimating  $f_t(x)$  have been widely used, and both involve regression splines. We will briefly describe them here.

## 2.1 Weighted constrained penalized regression splines

Hyndman & Ullah (2007) proposed using constrained and weighted penalized regression splines for estimating  $f_t(x)$ . The weighting takes care of the heterogeneity due to  $\sigma_t(x)$  and a monotonic constraint for upper ages can lead to better estimates.

Following Hyndman & Ullah (2007), we define weights equal to the approximate inverse variances  $w_{x,t} = m_{x,t}E_{x,t}$ , and use weighted penalized regression splines (Wood 2003, He & Ng 1999) to estimate the curve  $f_t(x)$  in each year. Weighted penalized regression splines are

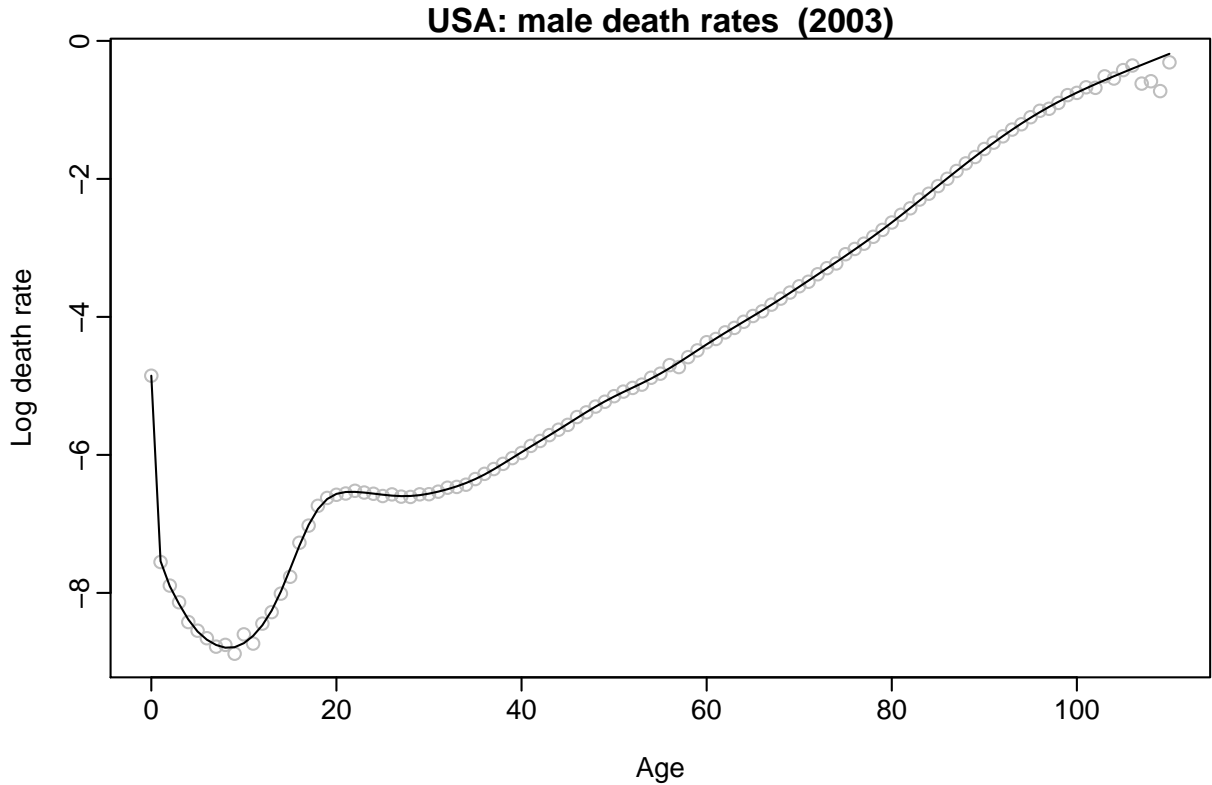


Figure 2: Smoothed male mortality rates for single years of age in the United States, 2003. The smooth curve,  $f_t(x)$ , is estimated using weighted penalized regression splines with a monotonicity constraint for ages greater than 65.

preferred because they can be computed quickly, and allow monotonicity constraints to be imposed relatively easily.

We apply a qualitative constraint to obtain better estimates of  $f_t(x)$ , especially when  $\sigma_t(x)$  is large. We assume that  $f_t(x)$  is monotonically increasing for  $x > c$  for some  $c$  (say 65 years). This monotonicity constraint allows us to avoid some of the noise in the estimated curves for high ages, and is not unreasonable for this application (after middle age, the older you are, the more likely you are to die). We use a modified version of the approach described in Wood (1994) to implement the monotonicity constraint.

Figure 2 shows the estimated smooth curve,  $f_t(x)$ , for the USA male mortality data plotted in Figure 1. This is easily implemented in the **demography** package in R using the following code.

```
smus <- smooth.demogdata(usa1950)
plot(usa1950, years=2003, series="male", type="p", pch=1, col="gray")
lines(smus, years=2003, series="male")
```

## 2.2 Two-dimensional P-splines

The above approach assumes  $f_t(x)$  is a smooth function of  $x$ , but not of  $t$ . Hyndman & Ullah (2007) argued that the occurrence of wars and epidemics meant that  $f_t(x)$  should not be assumed to be smooth over time. However, in the absence of wars and epidemics, it is reasonable to assume smoothness in both the time and age dimensions. Hence, Currie et al. (2004) proposed using two-dimensional splines instead. We will call this approach the Currie-Durban-Eilers or CDE method.

They adopt a generalized linear modelling (GLM) framework for the Poisson deaths  $D_{x,t}$  with two-dimensional B-splines. This is implemented in the **MortalitySmooth** package in R

(Camarda 2012) and compared with the Hyndman & Ullah (2007) approach using the following code.

```

74 Ext <- usa1950$pop$male
75 Dxt <- usa1950$rate$male * Ext
76 fitBIC <- Mort2Dsmooth(x=usa1950$age, y=usa1950$year, Z=Dxt, offset=log(Ext))
77
78 par(mfrow=c(1,2))
79
80 plot(fitBIC$x, log(usa1950$rate$male[, "2003"]), xlab="Year", ylab="Log death rate",
81      main="USA: male death rates 2003", col="gray")
82 lines(fitBIC$x, log(fitBIC$fitted.values[, "2003"]/Ext[, "2003"]))
83 lines(smus, year=2003, series="male", lty=2)
84 legend("topleft", lty=1:2, legend=c("CDE smoothing", "HU smoothing"))
85
86
87 plot(fitBIC$y, log(Dxt["65",]/Ext["65",]), xlab="Year", ylab="Log death rate",
88      main="USA: male death rates age 65", col="gray")
89 lines(fitBIC$y, log(fitBIC$fitted.values["65",]/Ext["65",]))
90 lines(smus$year, log(smus$rate$male["65",]), lty=2)
91 legend("bottomleft", lty=1:2, legend=c("CDE smoothing", "HU smoothing"))

```

Figure 3 shows the estimated smooth curve,  $f_t(x)$ , for the USA male mortality data using the bivariate P-spline method of Currie et al. (2004) and the univariate penalized regression spline method of Hyndman & Ullah (2007). Note that the univariate method is not smooth in the time dimension (right panel), but gives a better estimate for the oldest ages due to the monotonic constraint.

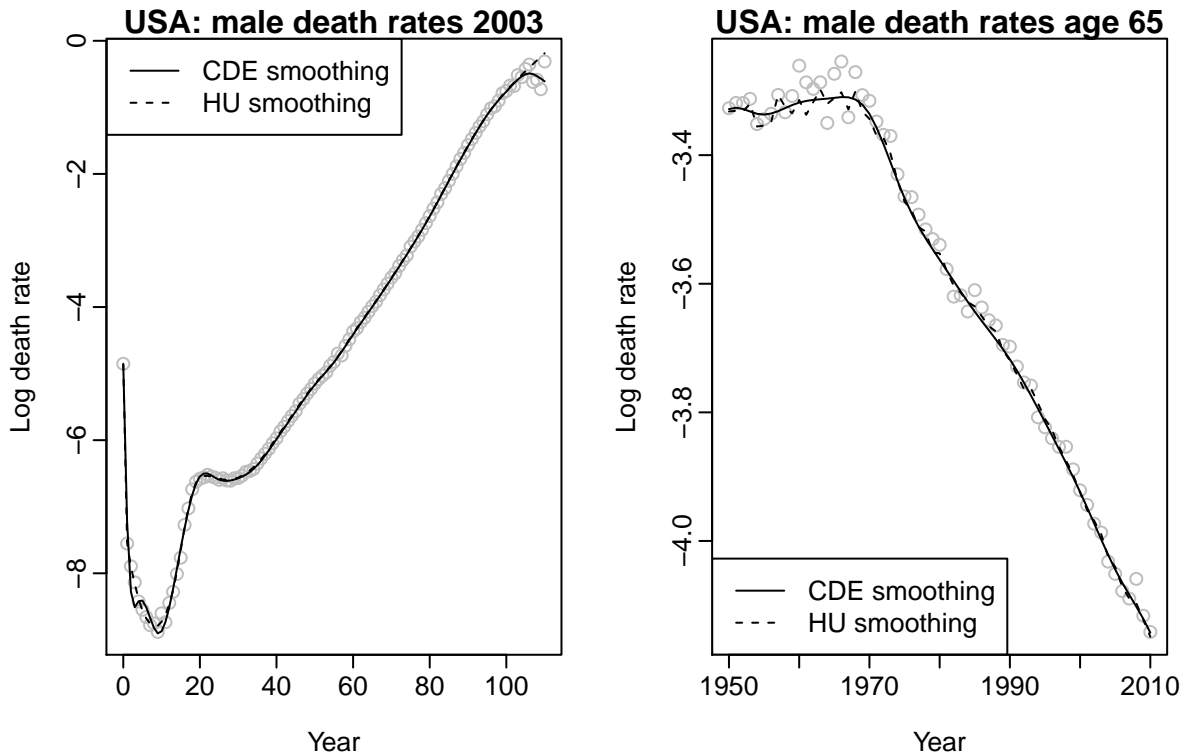


Figure 3: Smoothed male mortality rates using bivariate P-splines.

### 3 Lee-Carter and related forecasting methods

The Lee-Carter (LC) method (Lee & Carter 1992) for forecasting mortality rates uses principal components analysis to decompose the age-time matrix of central death (or mortality) rates into a linear combination of age and time parameters. The time parameter is used in forecasting.

LC has spawned numerous variants and extensions. The two main variants of LC are Lee-Miller (LM) (Lee & Miller 2001) and Booth-Maindonald-Smith (BMS) (Booth et al. 2002). Others result from different combinations of possible options. These variants are collectively referred to as “LC methods”. A major extension of this approach uses functional data analysis; first proposed by Hyndman & Ullah (2007), it was further developed by Hyndman & Booth (2008) and Hyndman & Shang (2009). Again, various combinations of options produce variations within the collectively labeled “HU methods”.

We identify six methods by their proponents; these are listed in Table 1 where the defining features of the models are shown. Most authors referring to the “Lee-Carter method” actually refer to the generic model in which all available data are used, there is no adjustment of the time parameter prior to forecasting, and fitted rates are used as jump-off rates; Booth et al. (2006) labeled this “LCnone”. Note that within the options listed in Table 1 there are 24 possible combinations (4 adjustment options  $\times$  3 data period options  $\times$  2 jump-off options) for the LC methods. For the HU methods, additional options have been defined by varying the data period option to include 1950 (Shang et al. 2011). Clearly, any date can be used for the start of the data period.

Method	Data period	Smoothing	Adjustment to match	Jump-off rates	Reference
Lee-Carter Methods					
LC	all	no	$D_t$	fitted	Lee & Carter (1992)
LM	1950	no	$e(0)$	actual	Lee & Miller (2001)
BMS	linear	no	$D_{x,t}$	fitted	Booth et al. (2002)
LCnone	all	no	—	fitted	—
Hyndman-Ullah Methods					
HU	all	yes	—	fitted	Hyndman & Ullah (2007)
HUrob	all	yes	—	fitted	Hyndman & Ullah (2007)
HUw	all	yes	—	fitted	Hyndman & Shang (2009)

Table 1: Lee-Carter and Hyndman-Ullah methods by defining features.

#### 3.1 Lee-Carter (LC) method

The model structure proposed by Lee & Carter (1992) is given by

$$\log(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}, \quad (1)$$

where  $a_x$  is the age pattern of the log mortality rates averaged across years;  $b_x$  is the first principal component reflecting relative change in the log mortality rate at each age;  $k_t$  is the first set of principal component scores by year  $t$  and measures the general level of the log mortality rates; and  $\varepsilon_{x,t}$  is the residual at age  $x$  and year  $t$ . The model assumes homoskedastic error and is estimated using a singular value decomposition.

The LC model in (1) is over-parameterized in the sense that the model structure is invariant under the following transformations:

$$\begin{aligned} \{a_x, b_x, k_t\} &\mapsto \{a_x, b_x/c, ck_t\}, \\ \{a_x, b_x, k_t\} &\mapsto \{a_x - cb_x, b_x, k_t + c\}. \end{aligned}$$

In order to ensure the model's identifiability, Lee & Carter (1992) imposed two constraints, given as:

$$\sum_{t=1}^n k_t = 0, \quad \sum_{x=x_1}^{x_p} b_x = 1.$$

In addition, the LC method adjusts  $k_t$  by refitting to the total number of deaths. This adjustment gives more weight to high rates, thus roughly counterbalancing the effect of using a log transformation of the mortality rates. The adjusted  $k_t$  is then extrapolated using ARIMA models. Lee & Carter (1992) used a random walk with drift (RWD) model, which can be expressed as:

$$k_t = k_{t-1} + d + e_t,$$

where  $d$  is known as the drift parameter and measures the average annual change in the series, and  $e_t$  is an uncorrelated error. It is notable that the RWD model provides satisfactory results in many cases (Tuljapurkar, Li & Boe 2000, Lee & Miller 2001, Lazar & Denuit 2009). From this forecast of the principal component scores, the forecast age-specific log mortality rates are obtained using the estimated age effects  $a_x$  and  $b_x$ , and setting  $\varepsilon_{x,t} = 0$ , in (1).

The LC method is implemented in the R **demography** package as follows:

```
lc.female <- lca(usa, series="female", ages=0:100)
forecast.lc.female <- forecast(lc.female, h=20)
```

The data (Figure 4), model parameters (Figure 5) and forecasts can be viewed via:

```
plot(usa, series="female")
plot(lc.female)
plot(forecast.lc.female, plot.type="component")
plot(usa, series="female", ylim=c(-10,0), lty=2)
lines(forecast.lc.female)
```

The LC method without adjustment of  $k_t$  (LCnone) is achieved by choosing the adjustment option "none".

```
lcnone.female <- lca(usa, series="female", adjust="none")
```

The effect of the LC adjustment of  $k_t$  is seen in Figure 6 via:

```
plot(lcnone.female$kt, ylab="kt", ylim=c(-70,90), xlab="")
lines(lc.female$kt, lty=2)
legend("topright", lty=1:2, legend=c("LCnone", "LC"))
```

An alternative, and more efficient, approach to estimating a Lee-Carter model was described by Brouhns et al. (2002), and involves embedding the method in a Poisson regression model, and using maximum likelihood estimation. This can be achieved in R using, for example,

```
lca(usa, series="female", adjust="dxt")
```

### 3.2 Lee-Miller (LM) method

The LM method is a variant of the LC method. It differs from the LC method in three ways:

1. the fitting period begins in 1950;
2. the adjustment of  $k_t$  involves fitting to the life expectancy  $e(0)$  in year  $t$ ;
3. the jump-off rates are the actual rates in the jump-off year instead of the fitted rates.

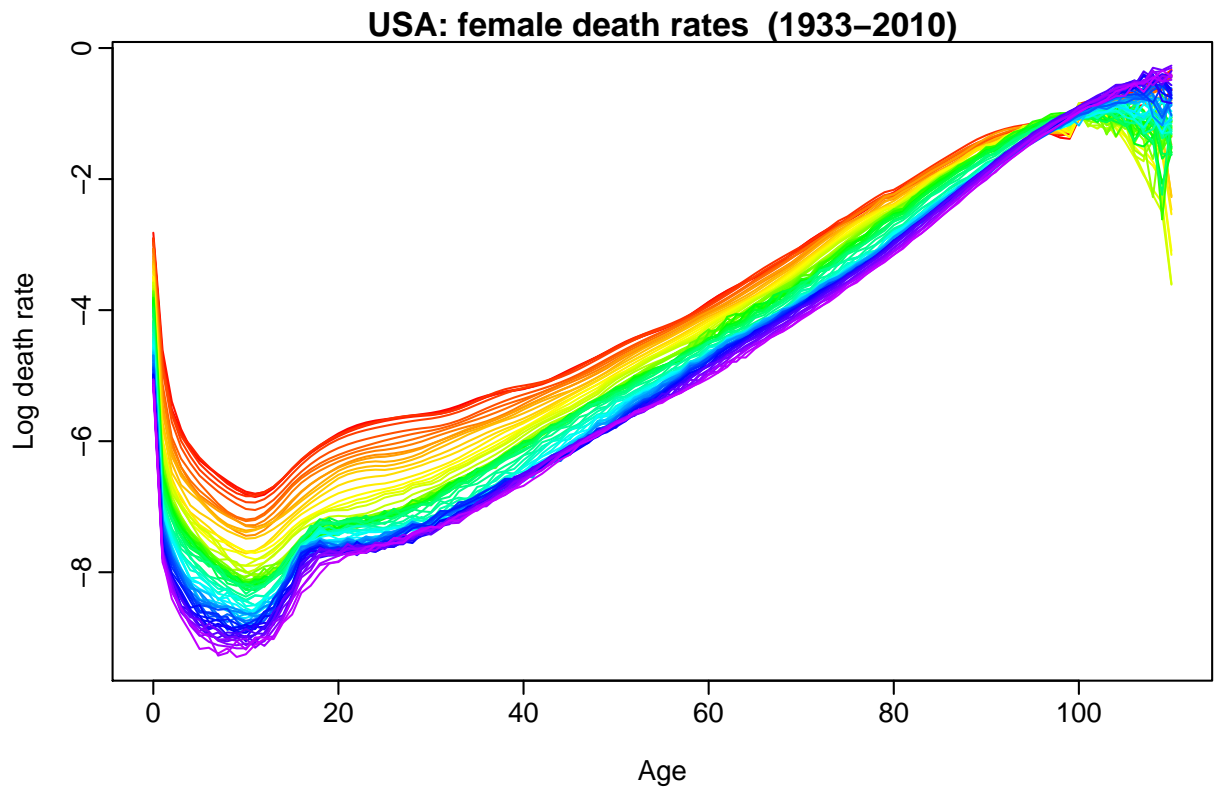


Figure 4: US female mortality rates, 1933–2010.

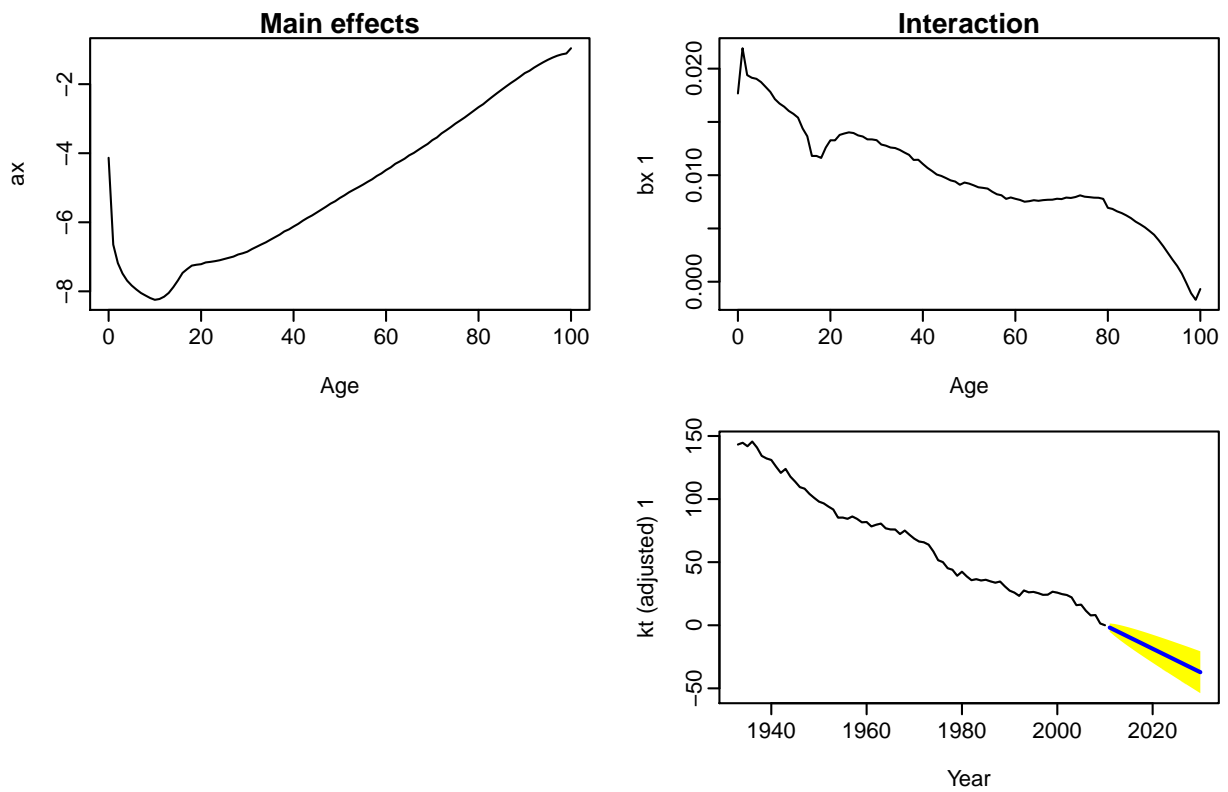


Figure 5: LC model and forecast, US female mortality. Fitting period = 1933–2010; forecasting horizon = 20 years.

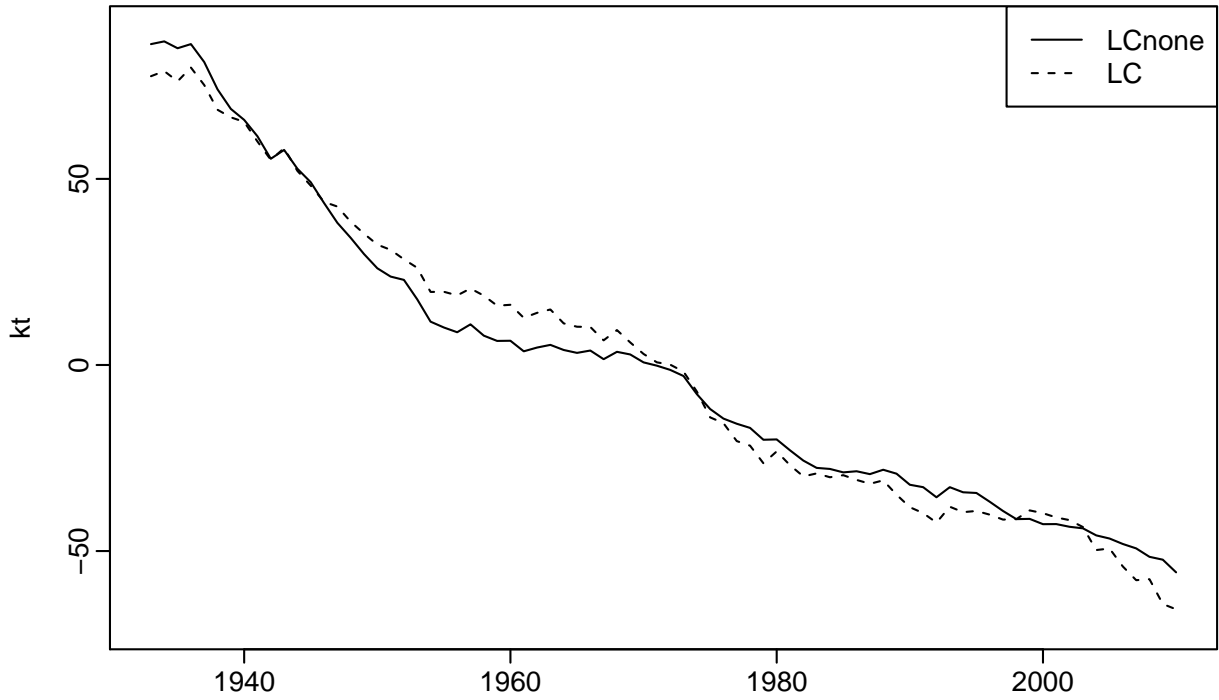


Figure 6: LC  $k_t$  with and without adjustment, US female mortality, 1933–2010.

In their evaluation of the LC method, Lee & Miller (2001) found that the pattern of change in mortality rates was not constant over time, which is a strong assumption of the LC method. Consequently, the adjustment of historical principal component scores resulted in a large estimation error. To overcome this, Lee & Miller (2001) adopted 1950 as the commencing year of the fitting period due to different age patterns of change for 1900–1949 and 1950–1995. This fitting period had previously been used by Tuljapurkar et al. (2000).

In addition, the adjustment of  $k_t$  was done by fitting to observed life expectancy in year  $t$ , rather than by fitting to total deaths in year  $t$ . This has the advantage of eliminating the need for population data. Further, Lee & Miller (2001) found a mismatch between fitted rates for the last year of the fitting period and actual rates in that year. This jump-off error was eliminated by using actual rates in the jump-off year.

The LM method is implemented as follows:

```
lm.female <- lca(usa, series="female", adjust="e0", years=1950:max(usa$year))
forecast.lm.female <- forecast(lm.female, h=20, jumpchoice = "actual")
```

The LM method has been found to produce more accurate forecasts than the original LC method (Booth et al. 2005, 2006).

### 3.3 Booth-Maindonald-Smith (BMS) method

The BMS method is another variant of the LC method. The BMS method differs from the LC method in three ways:

1. the fitting period is determined on the basis of a statistical ‘goodness of fit’ criterion, under the assumption that the principal component score  $k_1$  is linear;
2. the adjustment of  $k_t$  involves fitting to the age distribution of deaths rather than to the total number of deaths;
3. the jump-off rates are the fitted rates under this fitting regime.

A common feature of the LC method is the linearity of the best fitting time series model of the first principal component score, but Booth, Maindonald & Smith (2002) found the linear time series to be compromised by structural change. By first assuming the linearity of the



## Mean deviances for base and total models

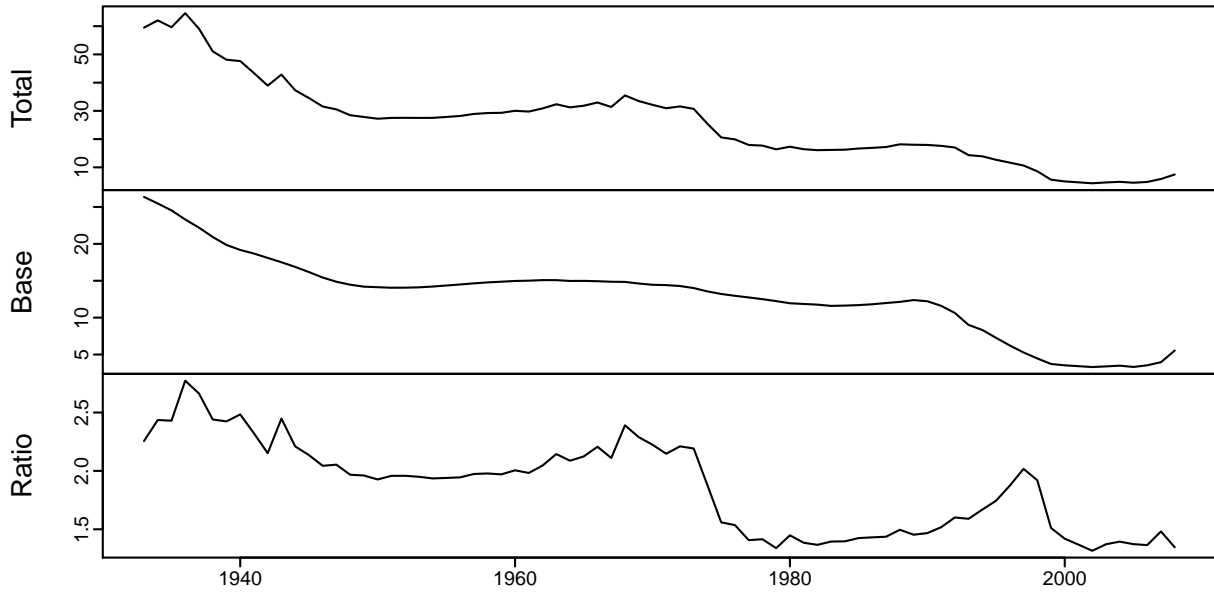


Figure 7: Mean deviances for base and total models and their ratio, US female mortality, 1933–2010.

first principal component score, the BMS method seeks to achieve the optimal ‘goodness of fit’ by selecting the optimal fitting period from all possible fitting periods ending in year  $n$ . The optimal fitting period is determined based on the smallest ratio of the mean deviances of the fit of the underlying LC model to the overall linear fit.

Instead of fitting to the total number of deaths, the BMS method uses a quasi-maximum likelihood approach by fitting the Poisson distribution to model age-specific deaths, and using deviance statistics to measure the ‘goodness of fit’ (Booth, Maindonald & Smith 2002). The jump-off rates are taken to be the fitted rates under this adjustment.

The BMS method is implemented thus:

```
bms.female <- bms(usa, series="female", minperiod = 20, breakmethod = "bms")
forecast.bms.female <- forecast(bms.female, h=20)
```

To view the deviances (Figure 7), chosen fitting period,  $k_t$  and forecast rates (Figure 8):

```
plot(bms.female$mdevs, main="Mean deviances for base and total models", xlab="")
bms.female$year[1]
plot(bms.female$kt)
plot(usa, series="female", ages=0:100, years=bms.female$year[1]:max(usa$year),
     ylim=c(-10,0), lty=2, main="Actual (1979-2010) and Forecast (2011-2030)")
lines(forecast.bms.female)
```

An alternative implementation using the `lca()` function, which permits all possible variants to be produced, is:

```
bms.female <- lca(usa, series="female", adjust="dxt", chooseperiod=TRUE,
                 minperiod = 20, breakmethod = "bms")
forecast.bms.female <- forecast(bms.female, h=20)
```

Forecasts from the BMS method have been found to be more accurate than those from the original LC method and of similar accuracy as those from the LM method (Booth et al. 2005, 2006).

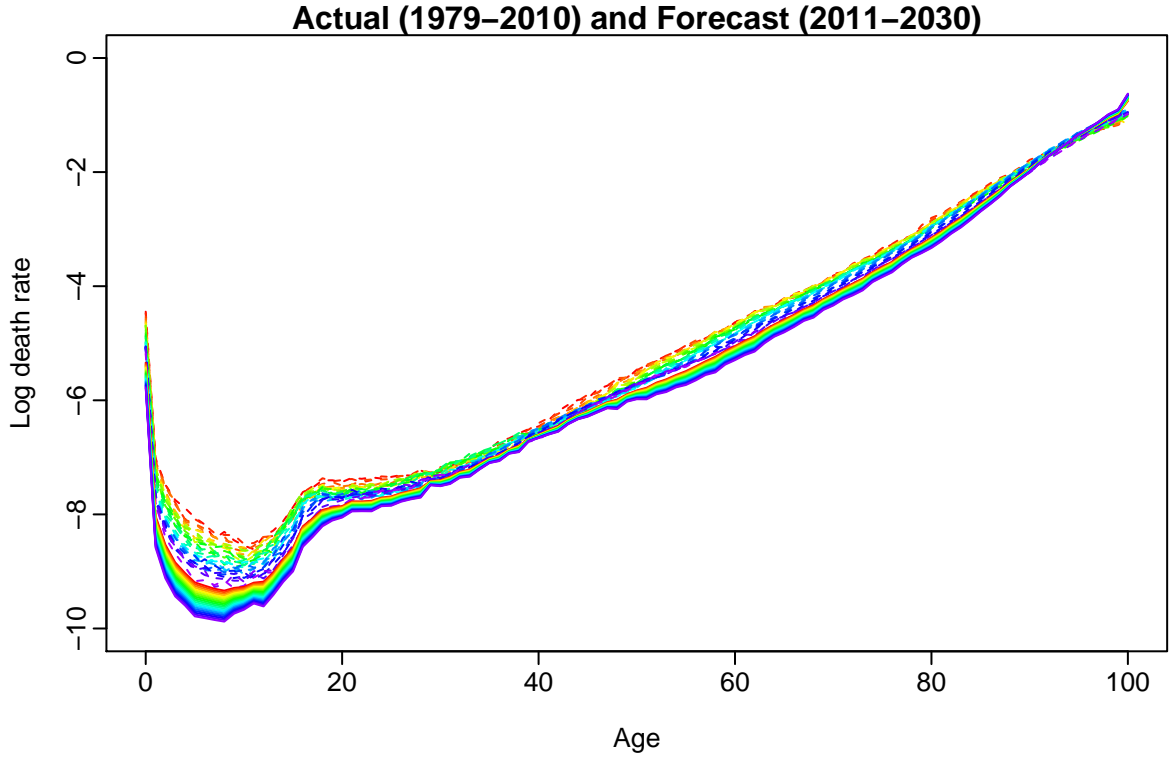


Figure 8: Actual (1979–2010) and forecast (2011–2030) mortality rates using the BMS method for US females.

### 3.4 Hyndman-Ullah (HU) method

Using the functional data analysis technique of Ramsay & Silverman (2005), Hyndman & Ullah (2007) proposed a nonparametric method for modeling and forecasting log mortality rates. This approach extends the LC method in four ways:

1. the log mortality rates are smoothed prior to modeling;
2. functional principal components analysis is used;
3. more than one principal component is used in forecasting;
4. the forecasting models for the principal component scores are typically more complex than the RWD model.

The log mortality rates are smoothed using penalized regression splines as described in Section 2. To emphasize that age,  $x$ , is now considered as a continuous variable, we write  $m_t(x)$  to represent mortality rates for age  $x \in [x_1, x_p]$  in year  $t$ . We then define  $z_t(x) = \log m_t(x)$  and write

$$z_t(x_i) = f_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}, \quad i = 1, \dots, p, \quad t = 1, \dots, n \quad (2)$$

where  $f_t(x_i)$  denotes a smooth function of  $x$  as before,  $\sigma_t(x_i)$  allows the amount of noise to vary with  $x_i$  in year  $t$ , thus rectifying the assumption of homoskedastic error in the LC model; and  $\varepsilon_{t,i}$  is an independent and identically distributed standard normal random variable.

Given continuous age,  $x$ , functional principal components analysis (FPCA) is used in the decomposition. The set of age-specific mortality curves is decomposed into orthogonal functional principal components and their uncorrelated principal component scores. That is,

$$f_t(x) = a(x) + \sum_{j=1}^J b_j(x)k_{t,j} + e_t(x), \quad (3)$$

where  $a(x)$  is the mean function estimated by  $\hat{a}(x) = \frac{1}{n} \sum_{t=1}^n f_t(x)$ ;  $\{b_1(x), \dots, b_J(x)\}$  is a set of the first  $J$  functional principal components;  $\{k_{t,1}, \dots, k_{t,J}\}$  is a set of uncorrelated principal

component scores;  $e_t(x)$  is the residual function with mean zero; and  $J < n$  is the number of principal components used. Note that we use  $a(x)$  rather than  $a_x$  to emphasise that  $x$  is not treated as a continuous variable.

Multiple principal components are used because the additional components capture non-random patterns that are not explained by the first principal component (Booth, Maindonald & Smith 2002, Renshaw & Haberman 2003, Koissi, Shapiro & Högnäs 2006). Hyndman & Ullah (2007) found  $J = 6$  to be larger than the number of components actually required to produce white noise residuals, and this is the default value. The conditions for the existence and uniqueness of  $k_{t,j}$  are discussed by Cardot, Ferraty & Sarda (2003).

Although Lee & Carter (1992) did not rule out the possibility of a more complex time series models for the  $k_t$  series, in practice a RWD model has typically been employed in the LC method. For higher order principal components, which are orthogonal by definition to the first component, other time series models arise for the principal component scores. For all components, the HU method selects the optimal time series model using standard model-selection procedures (e.g. AIC). By conditioning on the observed data  $\mathcal{I} = \{z_1(x), \dots, z_n(x)\}$  and the set of functional principal components  $\mathbf{B} = \{b_1(x), \dots, b_J(x)\}$ , the  $h$ -step-ahead forecast of  $z_{n+h}(x)$  can be obtained by:

$$\hat{z}_{n+h|n}(x) = E[z_{n+h}(x)|\mathcal{I}, \mathbf{B}] = \hat{a}(x) + \sum_{j=1}^J b_j(x) \hat{k}_{n+h|n,j},$$

where  $\hat{k}_{n+h|n,j}$  denotes the  $h$ -step-ahead forecast of  $k_{n+h,j}$  using a univariate time series model, such as the optimal ARIMA model selected by the automatic algorithm of Hyndman & Kandakdar (2008), or an exponential smoothing state space model (Hyndman et al. 2008).

Because of the orthogonality of all components, it is easy to derive the forecast variance as

$$\hat{v}_{n+h|n}(x) = \text{Var}[z_{n+h}(x)|\mathcal{I}, \mathbf{B}] = \sigma_a^2(x) + \sum_{j=1}^J b_j^2(x) u_{n+h|n,j} + v(x) + \sigma_t^2(x),$$

where  $\sigma_a^2$  is the variance of  $\hat{a}(x)$ ,  $u_{n+h|n,j}$  is the variance of  $k_{n+h,j} | k_{1,j}, \dots, k_{n,j}$  (obtained from the time series model),  $v(x)$  is the variance of  $e_t(x)$  and  $\sigma_t^2(x)$  is defined in (2). This expression is used to construct prediction intervals for future mortality rates in R.

The HU method is implemented as below. The model and forecast are seen in Figures 9 and 10.

```
fdm.male <- fdm(smus, series="male", order=3)
forecast.fdm.male <- forecast.fdm(fdm.male, h=30)
plot(forecast.fdm.male, plot.type="component")
plot(forecast.fdm.male)
```

### 3.5 Robust Hyndman-Ullah (HUrob) method

The presence of outliers can seriously affect the performance of modeling and forecasting. The HUrob method is designed to eliminate their effect. This method utilizes the reflection based principal component analysis (RAPCA) algorithm of Hubert, Rousseeuw & Verboven (2002) to obtain projection-pursuit estimates of principal components and their associated scores. The integrated squared error provides a measure of the accuracy of the principal component approximation for each year (Hyndman & Ullah 2007). Outlying years would result in a larger integrated squared error than the critical value obtained by assuming normality of  $e_t(x)$  (see Hyndman & Ullah 2007, for details). By assigning zero weight to outliers, the HU method can then be used to model and forecast mortality rates without possible influence of outliers.

The HUrob method is implemented as follows:

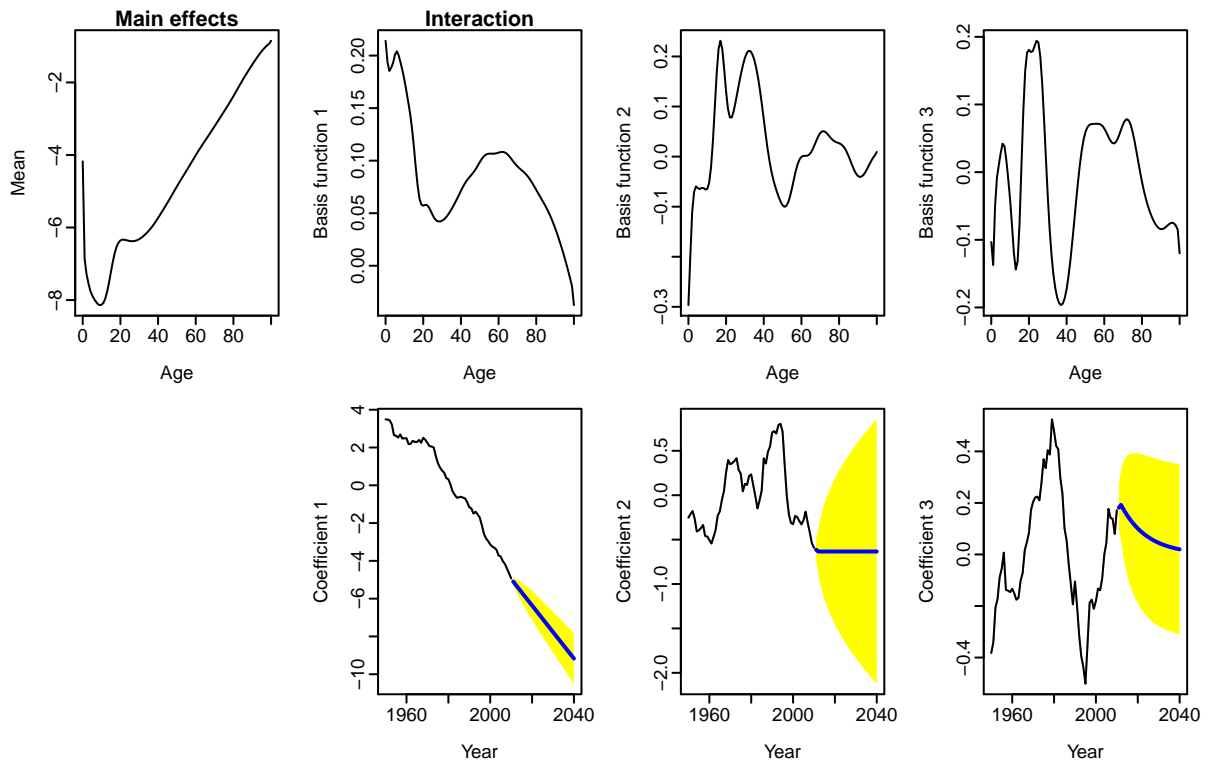


Figure 9: HU model and forecast, US male mortality. Fitting period = 1933-2010; forecasting horizon = 20 years.

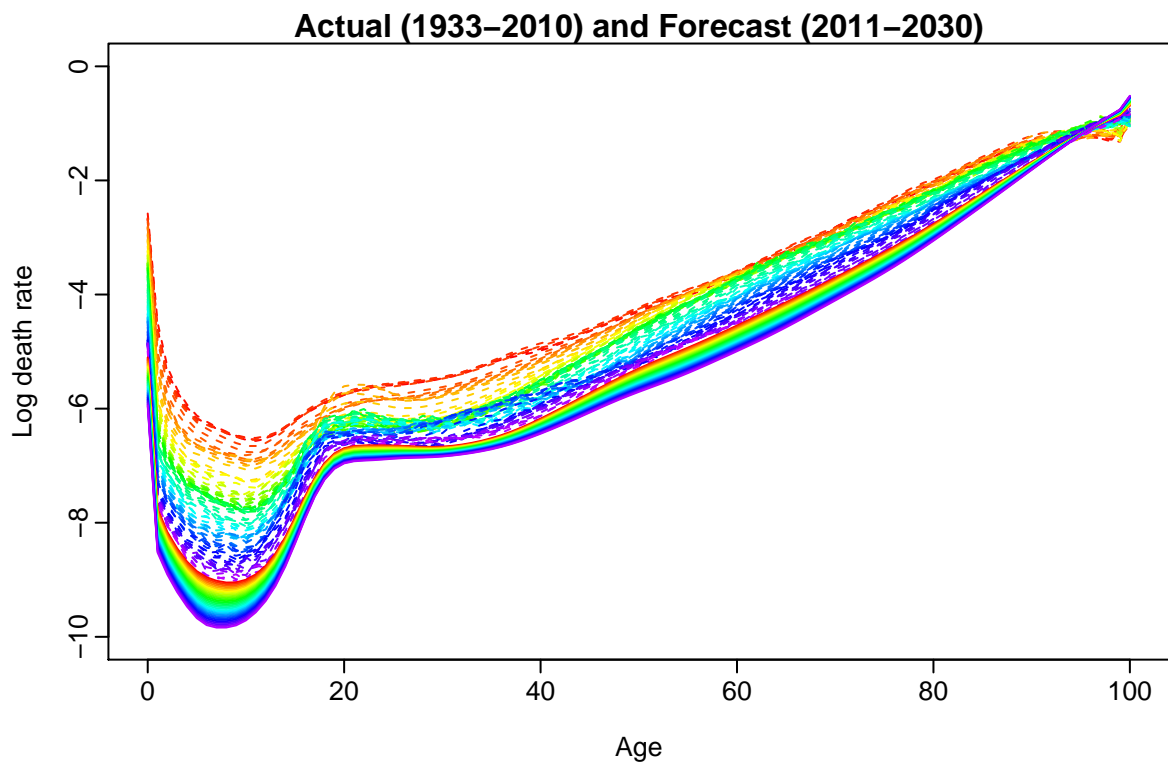


Figure 10: Actual and forecast mortality rates using the HU method for US males

```

260 fdm.male <- fdm(smus, series="male", method="rapca")
261 forecast.fdm.male <- forecast.fdm(fdm.male, h=20)
262 plot(forecast.fdm.male)

```

### 3.6 Weighted Hyndman-Ullah (HUw) method

The HU method does not weight annual mortality curves in the functional principal components analysis. However, it might be argued that more recent experience has greater relevance to the future than more distant experience. The HUw method uses geometrically decaying weights in the estimation of the functional principal components, thus allowing these quantities to be based more on recent data than on data from the distant past.

The weighted functional mean  $\hat{a}^*(x)$  is estimated by the weighted average

$$\hat{a}^*(x) = \sum_{t=1}^n w_t f_t(x), \quad (4)$$

where  $\{w_t = \beta(1 - \beta)^{n-t}, t = 1, \dots, n\}$  denotes a set of weights, and  $0 < \beta < 1$  denotes the weight parameter. Hyndman & Shang (2009) describe how to estimate  $\beta$  from the data. The set of weighted curves  $\{w_t[f_t(x) - \hat{a}^*(x)]; t = 1, \dots, n\}$  is decomposed using FPCA:

$$f_t(x) = \hat{a}^*(x) + \sum_{j=1}^J b_j^*(x) k_{t,j} + e_t(x), \quad (5)$$

where  $\{b_1^*(x), \dots, b_J^*(x)\}$  is a set of weighted functional principal components. By conditioning on the observed data  $\mathcal{I} = \{z_1(x), \dots, z_n(x)\}$  and the set of weighted functional principal components  $\mathbf{B}^*$ , the  $h$ -step-ahead forecast of  $z_{n+h}(x)$  can be obtained by:

$$\hat{z}_{n+h|n}(x) = E[z_{n+h}(x)|\mathcal{I}, \mathbf{B}^*] = \hat{a}^*(x) + \sum_{j=1}^J b_j^*(x) \hat{k}_{n+h|n,j}.$$

The HUw method is implemented as follows:

```
fdm.male <- fdm(smus, series="male", method="classical", weight=TRUE, beta=0.1)
forecast.fdm.male <- forecast.fdm(fdm.male, h=20)
plot(forecast.fdm.male)
```

## 4 Other mortality forecasting methods

Other extrapolative mortality forecasting methods are included here for completeness, but are not considered in detail as the methods are not fully implemented in packages available on CRAN.

A number of methods have been developed to account for the significant impact of cohort (year of birth) in some countries. In the U.K., males born around 1931 have experienced higher rates of mortality improvement than earlier or later cohorts (Willets 2004); less marked effects have also been observed elsewhere (Cairns et al. 2009).

The Renshaw and Haberman (RH) (2006) extension to Lee-Carter to include cohort effects can be written as<sup>1</sup>

$$\log(m_{x,t}) = a_x + b_x^1 k_t + b_x^2 \gamma_{t-x} + \varepsilon_{x,t}, \quad (6)$$

where  $a_x$  is the age pattern of the log mortality rates averaged across years,  $k_t$  represents the general level of mortality in year  $t$ ,  $\gamma_{t-x}$  represents the general level of mortality for the cohort born in year  $(t - x)$ ,  $b_x^1$  and  $b_x^2$  measure the relative response at age  $x$  to changes in  $k_t$  and  $\gamma_{t-x}$

<sup>1</sup>For clarity, models have been written in a standardised format which may in some cases differ from the form used by the authors originally.  $a_x$  and  $b_x$  terms are used for age-related effects,  $k_t$  terms for period-related effects, and  $\gamma_{t-x}$  terms for cohort-related effects. Models originally expressed in terms of the force of mortality are expressed in terms of the central mortality rate; these are equivalent under the assumption of a constant force of mortality over each year of age.

respectively, and  $\varepsilon_{x,t}$  is the residual at age  $x$ . The fitted  $k_t$  and  $\gamma_{t-x}$  parameters are forecast using univariate time series models. The model can be implemented using the **ilc** functions (Butt & Haberman 2009). The subsequent model

$$\log(m_{x,t}) = a_x + b_x k_t + \gamma_{t-x} + \varepsilon_{x,t} \quad (7)$$

(Haberman & Renshaw 2011), a special case of RH, was found to resolve some forecasting issues associated with the original. The Age-Period-Cohort (APC) model (Currie 2006), a special case of the revised model, incorporates age-, time- and cohort-effects that are independent in their effects on mortality.

$$\log(m_{x,t}) = a_x + k_t + \gamma_{t-x} + \varepsilon_{x,t} \quad (8)$$

The two-dimensional P-spline method of Currie et al. (2004) has already been described in Section 2. Forecast rates are estimated simultaneously with fitting the mortality surface. Implementation of the two-dimensional P-spline method to produce mortality forecasts use the **MortalitySmooth** package. Forecasts of USA male mortality rates and plots of age 65 and age 85 forecast rates with prediction intervals can be produced as follows, following the commands already shown in Section 2:

```
forecastyears <- 2011:2031
forecastdata <- list(x=usa1950$age, y=forecastyears)
CDEpredict <- predict(fitBIC, newdata=forecastdata, se.fit=TRUE)
whiA <- c(66,86)
plot(usa1950, series="male", age=whiA-1, plot.type="time",
     xlim=c(1950,2031), ylim=c(-6.2,-1), xlab="years",
     main="USA: male projected death rates using 2-dim CDE", col=c(1,2))
matlines(forecastyears, t(CDEpredict$fit[whiA,]), lty=1, lwd=2)
matlines(forecastyears, t(CDEpredict$fit[whiA,]+2*CDEpredict$se.fit[whiA,]), lty=2)
matlines(forecastyears, t(CDEpredict$fit[whiA,]-2*CDEpredict$se.fit[whiA,]), lty=2)
legend("bottomleft", lty=1, col=1:2, legend=c("Age65", "Age85"))
```

In addition to being applied in the age and period dimensions, two-dimensional P-spline method can incorporate cohort effects by instead being applied to age-cohort data.

Cairns et al. (2006) have forecast mortality at older ages using a number of models for  $\text{logit}(q_{x,t}) = \log[q_{x,t}/(1 - q_{x,t})]$ , where  $q_{x,t}$  is the probability that an individual aged  $x$  at time  $t$  will die before time  $t + 1$ . The original CBD model (Cairns et al. 2006) is

$$\text{logit}(q_{x,t}) = k_t^1 + (x - \bar{x})k_t^2 + \varepsilon_{x,t} \quad (9)$$

where  $\bar{x}$  is the mean age in the sample range. Later models (Cairns et al. 2009) incorporate a combination of cohort effects and a quadratic term for age:

$$\text{logit}(q_{x,t}) = k_t^1 + (x - \bar{x})k_t^2 + \gamma_{t-x} + \varepsilon_{x,t} \quad (10)$$

$$\text{logit}(q_{x,t}) = k_t^1 + (x - \bar{x})k_t^2 + ((x - \bar{x})^2 - \hat{\sigma}_x^2) + \gamma_{t-x} + \varepsilon_{x,t} \quad (11)$$

$$\text{logit}(q_{x,t}) = k_t^1 + (x - \bar{x})k_t^2 + (x_c - x)\gamma_{t-x} + \varepsilon_{x,t} \quad (12)$$

where the constant parameter  $x_c$  is to be estimated and the constant  $\hat{\sigma}_x^2$  is the mean of  $(x - \bar{x})^2$ . Other authors (e.g., Plat 2009) have proposed related models.

The LifeMetrics R software package implements the Lee-Carter method (using maximum likelihood estimation and a Poisson distribution for deaths) along with RH, APC,  $p$ -splines and the four CBD methods. The software, which is not part of CRAN, is available from [www.jpmmorgan.com/pages/jpmmorgan/investbk/solutions/lifemetrics/software](http://www.jpmmorgan.com/pages/jpmmorgan/investbk/solutions/lifemetrics/software). The software and the methods it implements is described in detail in Coughlan et al. (2007).

De Jong & Tickle (2006) (DJT) tailor the state space framework to create a method that integrates model estimation and forecasting, while using B-splines to reduce dimensionality and build in the expected smooth behaviour of mortality over age. Compared with Lee-Carter, the method uses fewer parameters, produces smooth forecast rates and offers the advantages of integrated estimation and forecasting. A multi-country evaluation of out-of-sample forecast performance found that LM, BMS, HU and DJT gave significantly more accurately forecast log mortality rates relative to the original LC, with no one method significantly more accurate than the others (Booth et al. 2006).

## 5 Coherent mortality forecasting

In modeling mortality for two or more sub-populations of a larger population simultaneously, it is usually desirable that the forecasts are non-divergent or “coherent”. The Product-Ratio method (Hyndman et al. 2013) achieves coherence through the convergence to a set of appropriate constants of forecast age-specific ratios of death rates for any two sub-populations. The method is an extension of functional forecasting (HU methods).

The method is presented here in terms of forecasting male and female age-specific death rates; extension to more than two sub-populations is straightforward (Hyndman et al. 2013). Let  $s_{t,F}(x) = \exp[f_{t,F}(x)]$  denote the smoothed female death rate for age  $x$  and year  $t$ ,  $t = 1, \dots, n$ . Similar notation applies for males.

Let the square roots of the products and ratios of the smoothed rates for each sex be

$$p_t(x) = \sqrt{s_{t,M}(x)s_{t,F}(x)} \quad \text{and} \quad r_t(x) = \sqrt{s_{t,M}(x)/s_{t,F}(x)},$$

These are modeled by functional time series models:

$$\log[p_t(x)] = \mu_p(x) + \sum_{k=1}^K \beta_{t,k} \phi_k(x) + e_t(x) \quad (13a)$$

$$\log[r_t(x)] = \mu_r(x) + \sum_{\ell=1}^L \gamma_{t,\ell} \psi_\ell(x) + w_t(x), \quad (13b)$$

where the functions  $\{\phi_k(x)\}$  and  $\{\psi_\ell(x)\}$  are the principal components obtained from decomposing  $\{p_t(x)\}$  and  $\{r_t(x)\}$  respectively, and  $\beta_{t,k}$  and  $\gamma_{t,\ell}$  are the corresponding principal component scores. The function  $\mu_p(x)$  is the mean of the set of curves  $\{p_t(x)\}$ , and  $\mu_r(x)$  is the mean of  $\{r_t(x)\}$ . The error terms, given by  $e_t(x)$  and  $w_t(x)$ , have zero mean and are serially uncorrelated.

The coefficients,  $\{\beta_{t,1}, \dots, \beta_{t,K}\}$  and  $\{\gamma_{t,1}, \dots, \gamma_{t,L}\}$ , are forecast using time series models as detailed in Section 3.4. To ensure the forecasts are coherent, the coefficients  $\{\gamma_{t,\ell}\}$  are constrained to be stationary processes. The forecast coefficients are then multiplied by the basis functions, resulting in forecasts of the curves  $p_t(x)$  and  $r_t(x)$  for future  $t$ . If  $p_{n+h|n}(x)$  and  $r_{n+h|n}(x)$  are  $h$ -step forecasts of the product and ratio functions respectively, then forecasts of the sex-specific death rates are obtained using  $s_{n+h|n,M}(x) = p_{n+h|n}(x)r_{n+h|n}(x)$  and  $s_{n+h|n,F}(x) = p_{n+h|n}(x)/r_{n+h|n}(x)$ .

The method makes use of the fact that the product and ratio behave roughly independently of each other provided the sub-populations have approximately equal variances. (If there are substantial differences in the variances, the forecasts remain unbiased but less efficient.)

The Product-Ratio method is illustrated in Figures 11, 12 and 13 and implemented as follows:

```
usa.pr <- coherentfdm(smus, weight=TRUE, beta=0.05)
usa.pr.f <- forecast(usa.pr, h=20)
plot(usa.pr.f$product, plot.type="component", components=3)
```

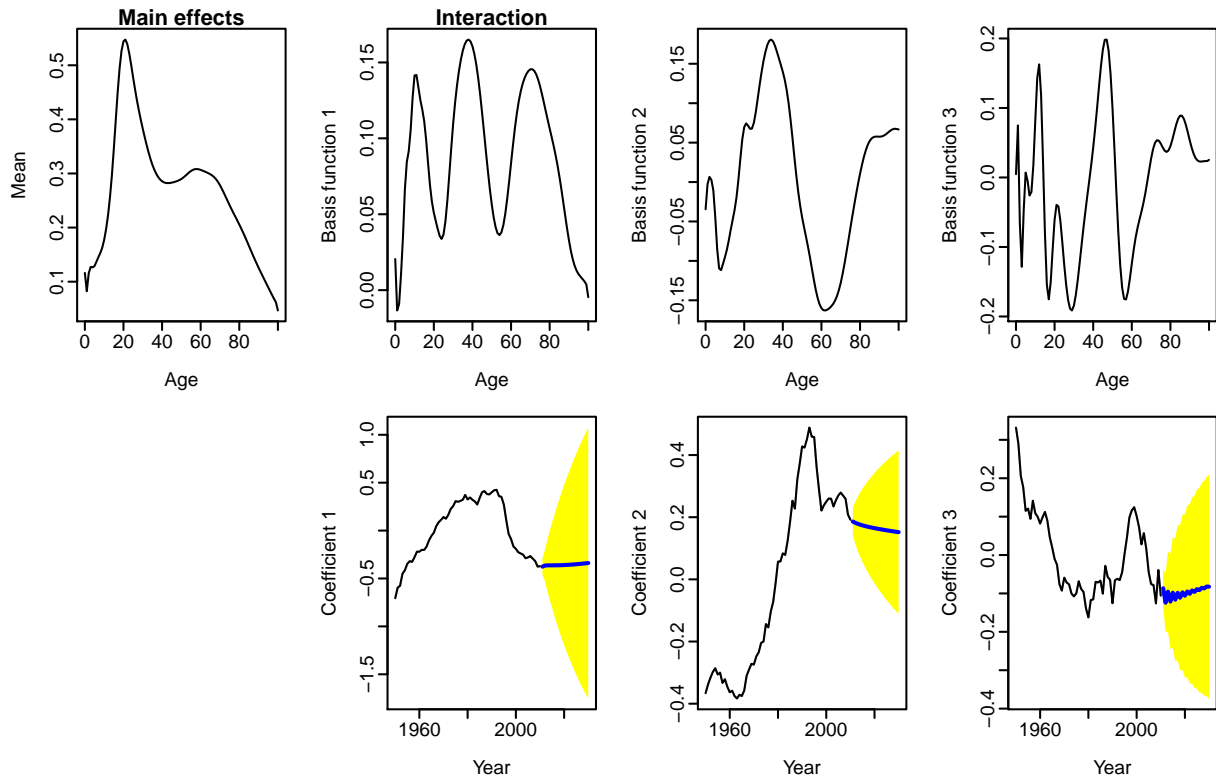


Figure 11: Ratio function decomposition with forecast

```

365 plot(usa.pr.f$ratio$male, plot.type="component", components=3)
366
367 par(mfrow=c(1,2))
368 plot(usa.pr$product$y, ylab="Log of geometric mean death rate", font.lab=2,
369 lty=2, las=1, ylim=c(-10,-1), main="Product function")
370 lines(usa.pr.f$product)
371 plot(sex.ratio(smus), ylab="Sex ratio of rates: M/F", ylim=c(0.7,3.5), lty=2,
372 las=1, font.lab=2, main="Ratio function")
373 lines(sex.ratio(usa.pr.f))
374
375 plot(smus, series="male", lty=2, ylim=c(-11,-1), main="Males")
376 lines(usa.pr.f$male)
377 plot(smus, series="female", lty=2, ylim=c(-11,-1), main="Females")
378 lines(usa.pr.f$female)

```

## 379 6 Life table forecasting

The methods described in this chapter generate forecast  $m_x$  rates, which can then be used to produce forecast life table functions using standard methods (e.g., Chiang 1984). Assuming that  $m_x$  rates are available for ages  $0, 1, \dots, \omega - 1, \omega +$ , the `lifetable` function in **demography**



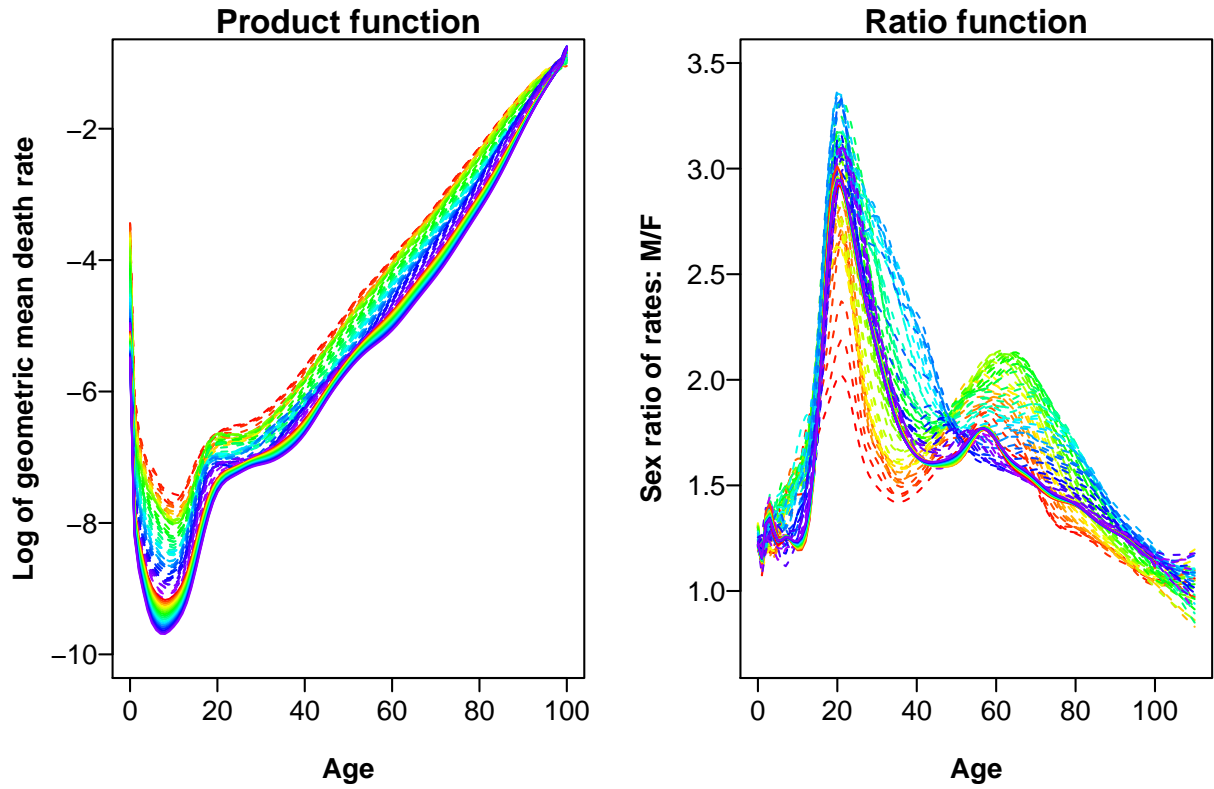


Figure 12: Actual(1933–2010) and forecast(2011–2030) product and ratio functions, USA.

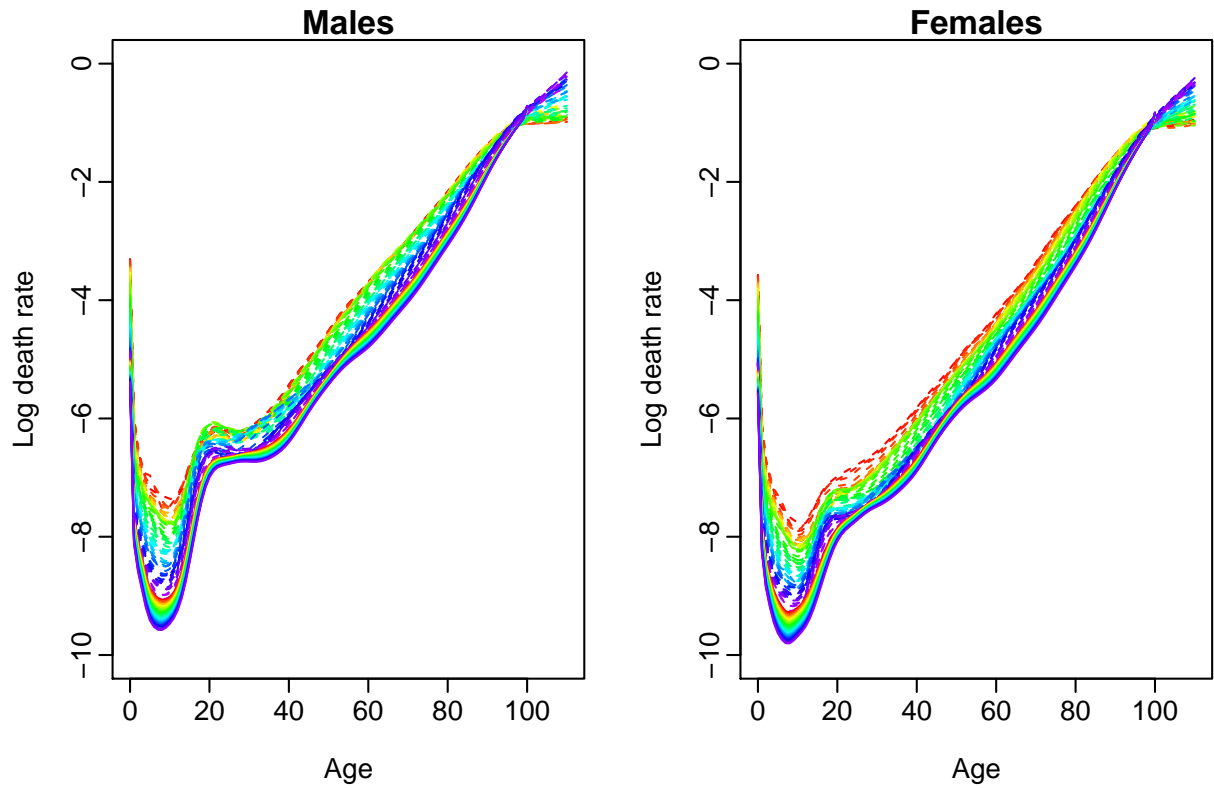


Figure 13: Actual(1933–2010) and forecast(2011–2030) male and female mortality rates using the product-ratio method with the FDM model, USA.

generates life table functions from a radix of  $l_0 = 1$  as follows for single years of age to  $\omega - 1$ :

$$q_x = m_x / (1 + (1 - a_x)m_x) \quad (14)$$

$$d_x = l_x q_x \quad (15)$$

$$l_{x+1} = l_x - d_x \quad (16)$$

$$L_x = l_x - d_x(1 - a_x) \quad (17)$$

$$T_x = L_x + L_{x+1} + \cdots + L_\omega \quad (18)$$

$$e_x = T_x / L_x \quad (19)$$

where  $a_x = 0.5$  for  $x = 1, \dots, \omega - 1$ , and  $a_0$  values (which allow for the fact that deaths in this age group occur earlier than midway through the year of age on average) are from Coale et al. (1983). For the final age group  $q_{\omega+} = 1$ ,  $L_{\omega+} = l_x / m_x$ , and  $T_{\omega+} = L_{\omega+}$ . For life tables commencing at an age other than zero, the same formulae apply, generated from a radix of 1 at the earliest age. The  $m_x$  rates on which the lifetable is based can be the rates applying in a future forecast year  $t$ , in which case a period or cross-sectional life table is generated, or can be rates that are forecast to apply to a certain cohort, in which case a cohort life table is generated.

The **demography** package produces life tables using `lifetable`, and life expectancies using the functions `e0`, `life.expectancy` and `flife.expectancy`. `flife.expectancy` is specifically designed for forecast life expectancies and will produce prediction intervals, and `e0` is a shorthand wrapper for `flife.expectancy` with `age=0`. All functions use the `cohort` argument to give cohort rather than period life tables and life expectancies.

To obtain prediction intervals for future life expectancies, we simulate the forecast log mortality rates as described in Hyndman & Booth (2008). In short, the simulated forecasts of log mortality rates are obtained by adding disturbances to the forecast basis function coefficients  $\beta_{t,k}$  and  $\gamma_{t,\ell}$  which are then multiplied by the fixed basis functions,  $\phi_k(x)$  and  $\psi_\ell(x)$ , respectively. Then, we calculate the life expectancy for each set of simulated log mortality rates. Prediction intervals are constructed from percentiles of the simulated life expectancies. This is all implemented in the **demography** package.

Using the coherent FDM model obtained in the previous section, we can forecast period life expectancies (Figure 14) as follows:

```
e0.fcast.m <- e0(usa.pr.f, PI=TRUE, series="male")
e0.fcast.f <- e0(usa.pr.f, PI=TRUE, series="female")
plot(e0.fcast.m, ylim=c(65,85), col="blue", fcol="blue")
par(new=TRUE)
plot(e0.fcast.f, ylim=c(65,85), col="red", fcol="red")
legend("topleft", lty=c(1,1), col=c("red", "blue"), legend=c("female", "male"))
```

An alternative approach to life expectancy forecasting is direct modeling, rather than via mortality forecasts. This is the approach taken by Raftery et al. (2013) who use a Bayesian hierarchical model for life expectancy, and pool information across countries in order to improve estimates. Their model is implemented in the **bayesLife** package, available on CRAN.

## 7 Exercises

1. Download the Human Mortality Database (HMD) mortality data for Denmark and plot male mortality rates at single ages 0 to 95+ for the 20th century.
2. Using data from 1950 for Danish females aged 0-100+, smooth the data by the Currie-Durban-Eilers and Hyndman-Ullah methods. Plot the two smoothed curves and the actual data for 1950 and 2000.

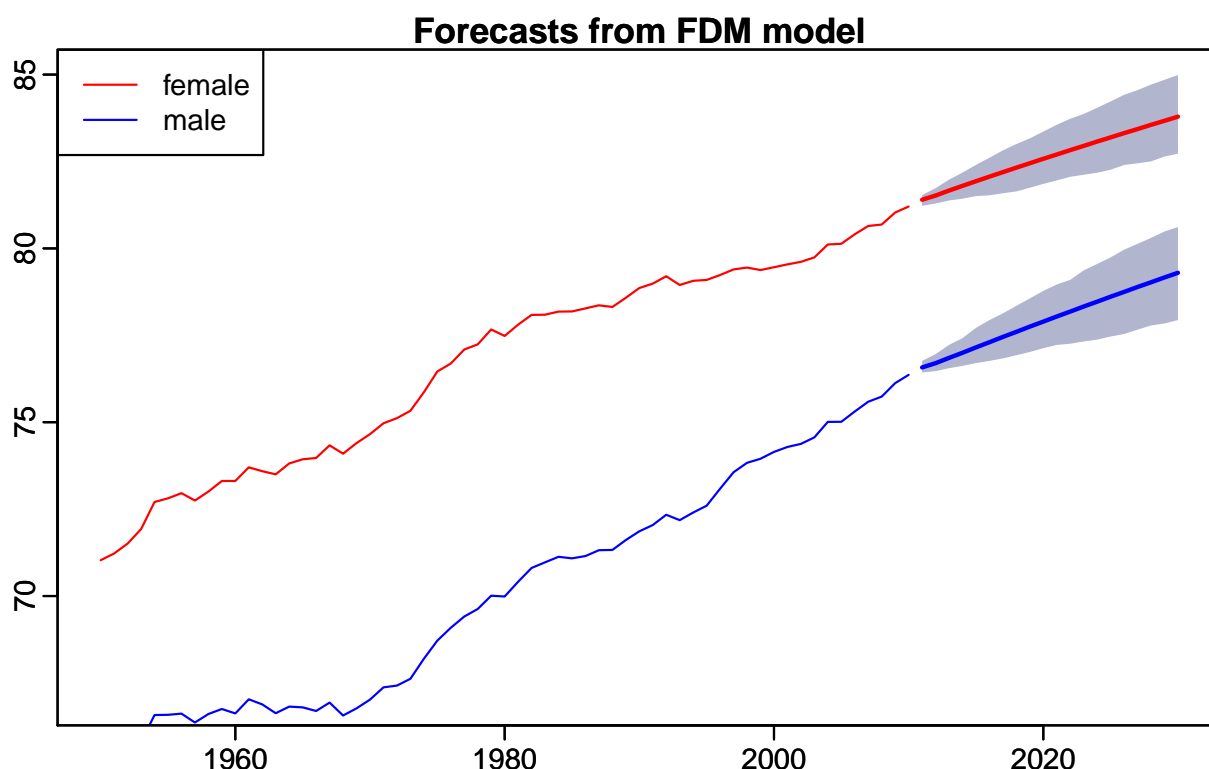


Figure 14: Period life expectancy forecasts for the United States calculated using the product-ratio method with the FDM model. Blue: males; red: females.

3. Download HMD data for Canada. Using data for the total population, compare forecast life expectancy for the next 20 years from the Lee-Carter and Lee-Miller methods.
4. Apply the Booth-Maindonald-Smith method to total mortality data for Canada. What is the fitting period? How does this forecast compare with the Lee-Miller forecast in terms of life expectancy after 20 years?
5. Using female data for Japan (from HMD), apply the Hyndman-Ullah method, and plot the first three components. Plot forecast mortality rates for the next 20 years. How does the forecast differ from a forecast of the same data using the Lee-carter method without adjustment?
6. Using male data for Japan, apply the Hyndman-Ullah method to forecast 20 years ahead, and plot male and female observed and forecast life expectancies on the same graph.
7. Apply the product-ratio method of coherent forecasting to data by sex for Japan. Plot past and future product and ratio functions. Add coherent male and female forecast life expectancies to the previous life expectancy graph.
8. Plot the sex difference over time in observed life expectancy, in independently forecast life expectancy and in coherently forecast life expectancy.

## References

- Booth, H. (2006), 'Demographic forecasting: 1980 to 2005 in review', *International Journal of Forecasting* **22**(3), 547–581.
- Booth, H., Hyndman, R. J., Tickle, L. & de Jong, P. (2006), 'Lee-Carter mortality forecasting : a multi-country comparison of variants and extensions', *Demographic Research* **15**(9), 289–310.

- Booth, H., Maindonald, J. & Smith, L. (2002), ‘Applying Lee-Carter under conditions of variable mortality decline’, *Population Studies* **56**(3), 325–336.
- Booth, H. & Tickle, L. (2008), ‘Mortality modelling and forecasting: A review of methods’, *Annals of Actuarial Science* **3**(1-2), 3–43.
- Booth, H., Tickle, L. & Smith, L. (2005), ‘Evaluation of the variants of the lee-carter method of forecasting mortality: a multi-country comparison’, *New Zealand Population Review* **31**(1), 13–37. Special Issue on Stochastic Population Projections, edited by A Dharmalingam & I Pool.
- Brillinger, D. R. (1986), ‘The Natural variability of vital rates and associated statistics’, *Biometrics* **42**, 693–734.
- Brouhns, N., Denuit, M. & Vermunt, J. K. (2002), ‘A Poisson log-bilinear regression approach to the construction of projected lifetables’, *Insurance: Mathematics and Economics* **31**(3), 373–393.
- Butt, Z. & Haberman, S. (2009), *ilc: A collection of R functions for fitting a class of Lee-Carter mortality models using iterative fitting algorithms*, Actuarial Research Paper 190, Cass Business School.  
**URL:** <http://goo.gl/oy5kK>
- Cairns, A. J. G., Blake, D. & Dowd, K. (2006), ‘A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration’, *Journal of Risk and Insurance* **73**(4), 687–718.
- Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A. & Balevich, I. (2009), ‘A quantitative comparison of stochastic mortality models using data from England and Wales and the United States’, *North American Actuarial Journal* **13**(1), 1–35.
- Camarda, C. G. (2012), ‘MortalitySmooth: An R package for smoothing Poisson counts with P-splines’, *Journal of Statistical Software* **50**(1), 1–24.  
**URL:** <http://www.jstatsoft.org/v50/i01>
- Cardot, H., Ferraty, F. & Sarda, P. (2003), ‘Spline estimators for the functional linear model’, *Statistica Sinica* **13**(3), 571–591.
- Chiang, C. L. (1984), *The Life Table and its Applications*, Robert E. Krieger Publishing.
- Coale, A., Demeny, P. & Vaughan, B. (1983), *Regional model life tables and stable populations*, Studies in population, Academic Press.
- Coughlan, G. D., Epstein, D., Ong, A., Sinha, A., Hevia-Portocarrero, J., Gingrich, E., Khalaf-Allah, M. & Joseph, P. (2007), *LifeMetrics: A toolkit for measuring and managing longevity and mortality risks*. Version 1.1: 13 March 2007. JP Morgan Pension Advisory Group.
- Currie, I. D. (2006), ‘Smoothing and forecasting mortality rates with P-splines’.
- Currie, I. D., Durbán, M. & Eilers, P. H. C. (2004), ‘Smoothing and forecasting mortality rates’, *Statistical Modelling* **4**(4), 279–298.
- De Jong, P. & Tickle, L. (2006), ‘Extending LeeCarter Mortality Forecasting’, *Mathematical Population Studies* **13**(1), 1–18.
- Haberman, S. & Renshaw, A. E. (2011), ‘A comparative study of parametric mortality projection models’, *Insurance: Mathematics and Economics* **48**(1), 35–55.

- He, X. & Ng, P. T. (1999), ‘COBS: qualitatively constrained smoothing via linear programming’, *Computational Statistics* **14**(3), 315–337.
- Hubert, M., Rousseeuw, P. J. & Verboven, S. (2002), ‘A fast method of robust principal components with applications to chemometrics’, *Chemometrics and Intelligent Laboratory Systems* **60**, 101–111.
- Human Mortality Database (2013), University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Downloaded on 22 February 2013.  
**URL:** *www.mortality.org*
- Hyndman, R. J. (2012), *demography: Forecasting mortality, fertility, migration and population data*. with contributions from Heather Booth, Leonie Tickle and John Maindonald.  
**URL:** *cran.r-project.org/package=demography*
- Hyndman, R. J. & Booth, H. (2008), ‘Stochastic population forecasts using functional data models for mortality, fertility and migration’, *International Journal of Forecasting* **24**(3), 323–342.
- Hyndman, R. J., Booth, H. & Yasmeeen, F. (2013), ‘Coherent mortality forecasting: the product-ratio method with functional time series models’, *Demography* **50**(1), 261–283.
- Hyndman, R. J. & Khandakar, Y. (2008), ‘Automatic time series forecasting : the forecast package for R’, *Journal of Statistical Software* **26**(3).
- Hyndman, R. J., Koehler, A. B., Ord, J. K. & Snyder, R. D. (2008), *Forecasting with exponential smoothing: the state space approach*, Springer-Verlag, Berlin.
- Hyndman, R. J. & Shang, H. L. (2009), ‘Forecasting functional time series (with discussion)’, *Journal of the Korean Statistical Society* **38**(3), 199–221.
- Hyndman, R. J. & Ullah, S. (2007), ‘Robust forecasting of mortality and fertility rates: A functional data approach’, *Computational Statistics and Data Analysis* **51**(10), 4942–4956.
- Koissi, M.-C., Shapiro, A. F. & Högnäs, G. (2006), ‘Evaluating and extending the Lee-Carter model for mortality forecasting: bootstrap confidence interval’, *Insurance: Mathematics and Economics* **38**(1), 1–20.
- Lazar, D. & Denuit, M. M. (2009), ‘A multivariate time series approach to projected life tables’, *Applied Stochastic Models in Business and Industry* **25**, 806–823.
- Lee, R. D. & Carter, L. R. (1992), ‘Modeling and forecasting U.S. mortality’, *Journal of the American Statistical Association* **87**(419), 659–671.
- Lee, R. D. & Miller, T. (2001), ‘Evaluating the performance of the Lee-Carter method for forecasting mortality.’, *Demography* **38**(4), 537–49.
- Plat, R. (2009), ‘On stochastic mortality modeling’, *Insurance: Mathematics and Economics* **45**(3), 393–404.
- Raftery, A. E., Chunn, J. L., Gerland, P. & Sevcíková, H. (2013), ‘Bayesian Probabilistic Projections of Life Expectancy for All Countries.’, *Demography* **50**, 777–801.
- Ramsay, J. O. & Silverman, B. W. (2005), *Functional Data Analysis*, 2nd edn, Springer-Verlag, New York.
- Renshaw, A. E. & Haberman, S. (2003), ‘Lee-Carter mortality forecasting with age-specific enhancement’, *Insurance: Mathematics and Economics* **33**(2), 255–272.

- 520 Renshaw, A. E. & Haberman, S. (2006), ‘A cohort-based extension to the Lee–Carter model for  
521 mortality reduction factors’, *Insurance: Mathematics and Economics* **38**(3), 556–570.
- 522 Shang, H. L., Booth, H. & Hyndman, R. J. (2011), ‘Point and interval forecasts of mortality  
523 rates and life expectancy : a comparison of ten principal component methods’, *Demographic  
524 Research* **25**(5), 173–214.
- 525 Tuljapurkar, S., Li, N. & Boe, C. (2000), ‘A universal pattern of mortality decline in the G7  
526 countries’, *Nature* **405**, 789–792.
- 527 Willets, R. C. (2004), *The cohort effect: insights and explanations*, Cambridge Univ Press.
- 528 Wood, S. N. (2003), ‘Thin plate regression splines’, *Journal of the Royal Statistical Society,  
529 Series B* **65**(1), 95–114.